Resources: lecture notes 3 and 4 (on Canvas), Kleinberg and Easley's *Networks, Crowds, and Markets* Chapter 2; Watts and Strogatz, Collective Dynamics of Small Worlds; Kleinberg, The Small World Phenomenon: An Algorithmic Perspective; section notes week 2 (on Canvas).

1. The square lattice. (20 points) In the following question we will consider a k-square lattice on n nodes. Recall that a 1-square lattice on n nodes with $r = \sqrt{n}$ can be thought of as a graph in which each node has some unique index (i, j) where $i, j \in \{1, \ldots, r\}$ and is connected to nodes with indices $(i-1, j), (i+1, j), (i, j-1), (i, j+1)^1$. The *lattice distance* between two nodes u and v is the number of edges between them on the 1-square lattice. A k-square lattice is a square lattice with the additional edges connecting every two nodes whose lattice distance is at most k. That is, a node u whose index on the lattice is (i, j) is connected to all nodes v whose indices are (x, y) where $|i - x| + |j - y| \leq k$. Consider such a k-square lattice on n nodes:

   a. (6 points) Compute its diameter for $k = 2$.

   b. (8 points) Consider a node u that is at lattice distance at least 2 from any nodes on the edge of the lattice (i.e. the node u has index (i, j) where $3 \leq i, j \leq r - 2$). Compute the clustering coefficient of node u (as defined in the previous problem set) for $k = 2$.

   c. (6 points) In the case of $k = 1$, and for even r, show that the expansion parameter of this graph is smaller than $2/r$. In other words, show that there exists a set of nodes S of size smaller or equal to $n/2$ (or $r^2/2$) has at most $|S| \cdot 2/r$ edges leaving it.

2. A smaller world? (Easley and Kleinberg, 20.8 Q1) (20 points) In the basic six-degrees-of-separation question, one asks whether most pairs of people in the world are connected by a path of at most six edges in the social network, where an edge joins any two people who know each other on a first-name basis. Now let's consider a variation on this question. Suppose that we consider the full population of the world, and suppose that from each person in the world we create an edge only to their ten closest friends (but not to anyone else they know on a first-name basis)^2. In the resulting "closest-friend" version of the social network, is it possible that for each pair of people in the world, there is a path of at most six edges connecting this pair of people? Explain.

3. Keep your friends close, and your acquaintances closer. (Easley and Kleinberg, 20.8 Q2) (20 points) We consider another variation of the six-degrees-of-separation question. For each person in the world, we ask them to rank the thirty people they know best, in descending order of how well they know them. (Let's suppose for purposes of this question that each person is able to think of thirty people to list.) We then construct two different social networks:

---

^{1}More precisely, (i, j) is connected to nodes $(i-1, j)$ if $i > 1$, $(i+1, j)$ if $i < r$, $(i, j-1)$ if $j > 1$, and $(i, j+1)$ $j < r$.

^{2}Assume that the relationship "closest friend" is symmetric, i.e. if Alice considers Bob to be one of her 10 closest friends, than Bob also considers Alice to be one of his 10 closest friends.

- The "close-friend" network: from each person we create a directed edge only to their ten closest friends on the list.

- The "distant-friend" network: from each person we create a directed edge only to the ten people listed in positions 21 through 30 on their list.

Let's think about how the small-world phenomenon might differ in these two networks. In particular, let C be the average number of people that a person can reach in six steps in the close-friend network, and let D be the average number of people that a person can reach in six steps in the distant-friend network (taking the average over all people in the world). When researchers have done empirical studies to compare these two types of networks (and the exact details often differ from one study to another), they tend to find that one of C or D is consistently larger than the other.

a. **(8 points)** For which of the close-friend network and the distant-friend network do you expect the clustering coefficient to be the largest?

b. **(8 points)** In which of the close-friend network and the distant-friend network do you expect to see the "long-range" edges that we talked about in the Watts-Strogatz and the Kleinberg models?

c. **(4 points)** Which of the two quantities, C or D, do you expect to be larger?

Give a brief explanation for each answer.

4. Regular graphs. (20 points) In this question you need to construct a graph that has certain properties. You can describe the graphs as a list, matrix, or drawing. We consider a graph with n nodes, and you can either be creative in describing a graph that generalizes to n nodes or simply use $n = 12$.

a. **(6 points)** Describe a connected 2-regular graph with diameter $n/2$ and clustering coefficient 0.

b. **(14 points)** Describe a connected 4-regular graph with diameter $n/4$ and clustering coefficient $1/2$.

5. Coding: short distances in networks. (20 points) Compute distances on networks

- (7 points) Recall the Watts-Strogatz model discussed in class. In this model, n nodes are positioned on a square lattice. Each node is connected to the nodes that are at most a distance k away on the lattice for some given parameter $k \in \mathbb{N}$. Moreover, each node has l *long range* edges which are connected uniformly at random to other nodes in the network. For $r = 1, 2, \ldots, 10$ (or $n = 1, 4, \ldots, 100$), generate networks according to the Watts-Strogatz model with $k = 2$ and $l = 2$. Plot the average distance between any two nodes as a function of n and plot the function $\log n$;

- (6 points) In the Kleinberg model, like in the Watts-Strogatz model, n nodes are positioned on a square lattice, and each node is connected to the nodes that are at most a distance k

away on the lattice for some given parameter $k \in \mathbb{N}$. Now, for some given l, each node has l long range edges that are generated as follows:

$$\Pr[u \to v] = \frac{\Delta(u,v)^{-2}}{\sum\limits_{w \neq u} \Delta(u,w)^{-2}}$$

$\Pr[u \to v]$ is the probability that the long range edge is created linking nodes u and v and $\Delta(u,v)$ is the lattice distance between nodes u and v.
For $n = 1, \ldots, 100$, generate networks according to the Kleinberg model with $k = 2$ and $l = 2$.
Plot the average distance between any two nodes as a function of n against $\log^2 n$;

- (7 points) Compute a **good estimate** (note: this does not need to be exact; clever solutions that provide good approximations of the shortest path are preferable to brute-force solutions) of the average shortest distances of these graphs:

  http://www.hcs.harvard.edu/~cs134-spring2017/wp-content/uploads/2017/01/enron.txt
  http://www.hcs.harvard.edu/~cs134-spring2017/wp-content/uploads/2017/01/epinions.
  txt
  https://www.dropbox.com/s/rzi61yf9f8rzgdv/livejournal.txt.zip?dl=0

  These files are formatted as lines of tab-separated values. For each line, n_1 n_2 corresponds to an edge from node n_1 to node n_2. In your solution, please include 1) your approximations of the average shortest distances of these graphs, 2) a brief written explanation describing your approach, 3) submit your code on Canvas. We have no specific style guidelines for your code, but please make it as easily-readable as you can.